The Construction and Validation of the OASys Violence Predictor: Advancing Violence Risk Assessment in the English and Welsh Correctional Services

Howard, Philip; Dixon, Louise

DOI:
10.1177/0093854811431239

Full reference for this post print article: Howard, P. D, & Dixon, L. (2012). The Construction and Validation of the OASys Violence Predictor: Advancing Violence Risk Assessment in the English and Welsh Correctional Services, *Criminal Justice and Behavior, 39, 287-307*

Journal homepage: http://cjb.sagepub.com/

The Construction and Validation of the OASys Violence Predictor: Advancing Violence Risk Assessment in the English and Welsh Correctional Services

By Philip D Howard[1] and Louise Dixon[2]

[1]National Offender Management Service, London, United Kingdom, and University of Birmingham

[2]University of Birmingham

Author Correspondence should be addressed to: Philip D. Howard, 26 Montana Close, South Croydon, Surrey, CR2 0AT United Kingdom. E-mail: philipdavidhoward@gmail.com

The Construction and Validation of the OASys Violence Predictor: Advancing Violence Risk Assessment in the English and Welsh Correctional Services

Abstract

The Offender Assessment System (OASys) is the risk assessment and management system routinely used in the National Offender Management Service (NOMS), the prison and probation service for England and Wales. This study describes the construction and validation of a new actuarial violence risk measure, the OASys Violence Predictor (OVP), using OASys and Police National Computer data. Ordinal logistic regression identifies static and dynamic risk factors predictive of violent recidivism among convicted offenders (N = 15,918). These form the basis of a user-friendly 100-point scale (OVP). OVP achieves significantly greater predictive validity than existing actuarial scores available within NOMS (the original OASys risk prediction score, OGRS and Risk Matrix 2000/V) on a later validation sample (N = 49,346). The discussion considers explanations for this improvement, examines the utility of dynamic risk factors in violence prediction, and describes the application of OVP in NOMS's treatment allocation and risk management practice.

Keywords: OASys; violence risk assessment; UK correctional services; dynamic risk factors; AUC.

Introduction

Empirical developments in forensic risk assessment have been pivotal in increasing the accuracy of decisions about the likely risk an offender poses to the public (Gottfredson & Moriarty, 2006) and identifying criminogenic factors for intervention (Andrews & Bonta, 2006). A range of risk assessment tools for violent reoffending is now available (Hanson, 2005; Heilbrun, Yasuhara & Shah, 2009). However, research examining the role of dynamic risk factors (Douglas & Skeem, 2005; Brown, St. Amand & Zamble, 2009) and proposals for new measures and approaches (Edens & Douglas, 2006; Walters, 2007; Howells, 2009) continue to be developed and debated.

Some researchers question whether it is possible to produce new, more accurate, predictors of violent recidivism. According to this view, inherent unpredictability in offending behaviour and the criminal justice system impose a “ceiling” on predictive validity which may now have been reached (Campbell, French & Gendreau, 2007; Yang, Wong & Coid 2010). High levels of intercorrelation and content overlap are found among predictors of violent reoffending (Campbell et al, 2007; Yang et al, 2010; Kroner, Mills & Reddon, 2005) and actuarial predictors of sexual reoffending. In some circumstances, there is however a strong case for the development of a new violence risk assessment instrument. This manuscript describes the construction and validation of such a tool for the National Offender Management Service (NOMS) of England and Wales. NOMS is responsible for the supervision of offenders aged 18 and over serving sentences in custody and the community, numbering 83,500 (June 2009) and 174,000 (December 2009) respectively (Ministry of Justice, 2010a). NOMS’s primary risk/needs assessment instrument is the Offender Assessment System (OASys), a tool which has seldom been studied in forensic psychology and is described in detail in the Materials section of this manuscript.

OASys has been broadly successful in providing an effective structure for risk assessment with which probation staff is comfortable (Mair, Burke & Taylor, 2006). NOMS is committed to staff development and quality assurance procedures which aim to ensure that offender management practice addresses the risks and needs identified in assessments (Ministry of Justice, 2006). However, significant problems relating to predictive validity have been evident. The original OASys scoring system was shown to be an inadequate predictor of general reoffending (Howard, 2006) and, in subsequent unpublished analysis, violent reoffending. NOMS policymakers mandated that the scoring system should be revised, with particular attention paid to the prediction of harmful recidivism. Valid prediction of harmful recidivism is necessary in order to correctly assess future risks when reporting to courts (National Probation Service, 2009), and the Parole Board (HM Prison Service, 2009), determining intensity of probation supervision (National Probation Service, 2008), and allocating places on offending behavior programmes (Palmer, McGuire, Hatcher, Hounsome, Bilby & Hollin, 2009). Instruments specifically designed to assess the risk of violent recidivism, including HCR-20 (Webster, Douglas, Eaves & Hart, 1997) and VRAG (Quinsey, Harris, Rice & Cormier, 1998), are already used to assess some offenders in the NOMS caseload, such as those serving lengthy prison sentences. However, NOMS policymakers determined that the additional time required to complete these assessments precluded their general introduction, given the large caseloads handled by NOMS staff (Ministry of Justice, 2010a), and intensifying budgetary pressures (HM Treasury, 2011). The only practical options for widespread violence risk prediction in the NOMS caseload would be the use of rapidly scored static actuarial tools – OGRS3 (Howard, Francis, Soothill & Humphries, 2009) or the 'V' scale of RM2000 (Thornton, 2007), described below – or the development of a predictor within OASys, as its continued use by NOMS is certain for the foreseeable future. The decision was therefore taken within NOMS that an OASys-based violence predictor should be developed. Its predictive validity could then be compared with the original OASys score, OGRS3 and RM2000/V, and a decision taken on whether to implement it as a revision to the operational OASys system.

Existing literature offered some evidence for the likely success of an OASys-based predictor. Firstly, scales from assessment tools designed on risk/needs principles can predict violence well: Campbell et al (2007) and Yang et al (2010) found that LSI-R (Andrews & Bonta, 1995) had similar validity to several other tools. Campbell et al also found that LS/CMI (Andrews, Bonta & Wormith, 2004) had very strong effect sizes in a small number of studies. Secondly, this study develops a predictor using a mixture of OASys items, many of which are scored through structured professional judgement, and static actuarial items such as age and criminal history; Douglas, Yeomans and Boer (2005) found that HCR-20's SPJ ratings and VRAG's static actuarial risk bins made independent contributions to predictive validity.

Aims and Requirements in the Development of a NOMS Predictor of Violent Offending

The new predictor, named the OASys Violence Predictor (OVP), is intended to be easy to use, while rigorously combining static and dynamic risk factors. It must support key NOMS requirements, by being quick and simple enough to use as part of its day-to-day processes such as case prioritisation and intervention targeting, reflecting changes in risk over the course of a sentence, and upholding its public protection objective (Ministry of Justice, 2006).

Pressures on the time spent assessing and managing each offender make it imperative that changes to OASys do not increase its complexity. OVP's criminal history items should be simple enough to be calculated reliably by the auxiliary staff who often complete OASys's static risk components. The scoring system should allow offender managers to understand the association between an individual offender's risk factors and their OVP score. OASys users were consulted to ensure that OVP fulfilled these practical requirements, as described later.

As OASys assessments are administered repeatedly over the course of a sentence, offenders' scores include dynamic risk factors that will change over time. The potential for significant change improves OVP's value in processes which typically occur late in the sentence, including parole hearings and evaluation of progress on interventions designed to address offending behaviour. The integrity and defensibility of the parole process requires that risk assessment instruments used must utilise offenders' individual characteristics at the time of the parole hearing, maximise predictive validity, identify risk of recidivism at different points in time, be relevant to male and female offenders and enable assessors to determine the risk of onset and persistence of violent offending.

NOMS's public protection objective requires the identification of offenders likely to commit those violent offences which cause the most serious harm. As the sexual recidivism risk of all known sex offenders is already assessed using Risk Matrix 2000/S (Thornton, 2007) and, after treatment, the Structured Assessment of Risk and Need (SARN; Thornton, 2002), OVP focuses on valid estimation of serious nonsexual violent reoffending risk.

Four steps of analysis were conducted to construct and validate OVP in accordance with these three key requirements:

- First, a logistic regression model of violent reoffending was constructed from static and dynamic risk factors.
- Second, a simplified scoring system (OVP) was developed from the regression model results, and the impact of the simplification process on predictive validity was tested using a separate validation sample of assessments completed in a later time period.
- Third, the predictive validities of OVP and other actuarial risk prediction instruments used routinely in NOMS (OGRS and RM2000/V; see Measures), were compared, using the validation sample. OGRS and RM2000/V are NOMS's only realistic alternatives to OVP. These actuarial tools satisfy the simplicity requirement, but are not dynamic.

- Fourth, the predictive validity of OVP for key subgroups – male and female offenders, and those with and without a history of violent offending – was checked, using the validation sample. Comparison with the validation sample signals whether this approach produced any worthwhile improvement in predictive validity.

In addition, OASys users were consulted to ensure that OVP met NOMS's usability requirements.

Method

Measures

Offender Assessment System (OASys).

The Offender Assessment System (OASys) (Home Office, 2006) is a structured clinical risk/needs assessment and management tool constructed on risk/need/responsivity principles. It is used throughout NOMS. Before sentence, it is used to inform court reports on convicted offenders. Later, it is used to help manage offenders serving custodial sentences of at least 12 months (which are usually partially served in the community) or community sentences involving supervision. Assessments are reviewed periodically over the course of the sentence. In 2008/09, approximately 830,000 assessments were completed on 360,000 offenders by 12,000 staff. All staff are trained in offending behaviour theories, assessment-related skills (e.g., interviewing offenders, clinical case formulation) and the use of OASys itself. All assessments are countersigned by a more senior officer, and samples of assessments are scrutinised more thoroughly in quality assurance exercises. OASys has strongly influenced the design of the offender assessment systems of several other European countries (van Kalmthout & Durnescu, 2008).

OASys consists of four main components: an analysis of offending-related factors, a risk of serious harm analysis, a summary sheet and a sentence plan. The offending-related factors component includes 13 sections, covering criminal history, Analysis of [current] Offences, assessment of ten dynamic risk factors (this study's dynamic factors: see Table 2) and suitability to undertake sentence-related activities (e.g., unpaid work, offending behavior programs). Each dynamic risk factor is assessed using between four and ten questions, each scored on a 0/2 or 0/1/2 basis. The risk of serious harm analysis (RoSH) component provides a structure for clinical case formulation and a Risk Management Plan for offenders considered likely to commit harmful acts in the future. The summary sheet component automatically scores predictors of recidivism using IT functionality. Prior to the implementation of OVP, the summary sheet calculated a total "OASys score". This combined scores from each of the offending-related factors, with item weights set by OASys's original designers before reconviction studies had been conducted. Howard (2006) showed that the OASys score had only moderate predictive validity. The summary sheet now instead presents scores on OVP and the complementary OASys General reoffending Predictor (OGP), a predictor of nonviolent reoffending described in Howard (2009). OVP scores are also presented to assessors during the RoSH. The Sentence Plan assesses responsivity considerations, which are then combined with the dynamic risk factors and RoSH to determine case management strategies and interventions. These should fulfil sentence requirements, manage the offender's risk of serious harm and reduce their likelihood of reoffending. Violent reoffending is therefore the focus of many Risk Management and Sentence Plans, as well as reports provided to decision makers such as sentencers and the Parole Board. Despite this, until the implementation of OVP, OASys did not specifically assess offenders' risk of future violent offending.

Moore (2009) examined the internal reliability and construct validity of the ten dynamic risk factor sections and the criminal history section. Eight of these sections were described by single factors, but three split into two factors each and a further 'violence' factor emerged. Morton (2009) produced promising but methodologically weak inter-rater reliability results.

Howard and Moore (2009) compared item and section (risk factor) scores over series of assessments during community supervision periods of up to two years. They found that many of OASys's risk factors are dynamic in several key respects. Most item scores changed in between 5% and 20% of original/final assessment pairs, only 30% of such assessment pairs included no changes in any dynamic item score, and changes in section scores between first and second assessments were predictive of recidivism at third assessment.

The Offender Group Reconviction Scale (OGRS3).

OGRS is used slightly more frequently than OASys, as it is also used for oral court reports and nonrehabilitative sentences such as Community Orders involving unpaid work. It is a purely actuarial calculation of the probability of proven reoffending for most recordable offences, combining criminal history and demographic variables in a logistic function. It has been periodically revised and recalibrated, and version 3 (OGRS3; Howard et al, 2009) has recently been introduced. OGRS achieved a weighted, adjusted Area Under Curve (AUC) of .71 from two violence prediction studies in Yang et 's (2010) meta-analysis.

Risk Matrix 2000/V (RM2000).

RM2000 (Thornton, 2007) is an actuarial predictor of the likelihoods of reconviction for sexual offences ('S' scale), nonsexual violent offences ('V' scale) and either of these offence groups ('C' scale). While it was constructed to predict for adult male sex offenders, the 'V' scale has been successfully validated on a sample of prisoners without histories of sexual offending (Thornton, 2007). It is scored from three factors: age, violent appearances and having any convictions for burglary. This yields a score from 0 to 8, which is banded into four risk categories. RM2000/V has seldom been included in validation studies not exclusively focused on sex offenders, but obtained a weighted, adjusted AUC of .69 from three studies in Yang et al's (2010) meta-analysis.

Previous sanctions and proven reoffending.

Previous sanctions for an offence group are the number of formal criminal sanctions (convictions, cautions, reprimands and final warnings) the offender received for that offence group up to and including the index sanction. Proven reoffending comprises offences committed within 24 months of the date of community sentence or release from custody, leading to formal criminal sanction no more than 12 months after the end of this followup period.

Proven violent reoffending is classified in OVP as any proven reoffending involving offence(s) of homicide and assault, threats and harassment, violent acquisitive offences (robbery and aggravated burglary), public order, criminal damage and/or weapon possession. These offences are coded from lists maintained by the UK Home Office. Howard and Dixon (2010) examined patterns of index offence content and dynamic risk factors in 230,000 OASys assessments, and associations between previous violent sanctions, dynamic risk factors and recidivism in the present study's 2002-04 data sample. They determined that this classification of violent offending was likely to aid prediction of future homicide and assault, including a subset of the most serious offences named “homicide and wounding” (i.e., murder, attempted murder, nonvehicular manslaughter and grievous bodily harm with intent). Contact sexual offences were excluded from this violent offence classification as they were shown to be unlikely to aid such prediction. To reiterate Howard and Dixon’s point, this does not imply that sexual offences are not harmful. OVP uses this classification to determine which offences count as previous violent and nonviolent sanctions, and which count as proven violent reoffending. The validation comparisons checked prediction of violent but also homicide/assault and homicide/wounding proven reoffending outcomes. These additional checks ensure that OVP does improve prediction of these more serious subsets of offences.

Procedure

The Police National Computer (PNC) research database.

The Police National Computer (PNC) is the operational system used by all 42 police forces in England and Wales. It records details of suspected and proven offenders, as well as details of crimes solved and under investigation. The Ministry of Justice's PNC research database (MoJPNC) contains extracts of PNC criminal records data on cautioned and convicted offenders. It is available to researchers through the Ministry of Justice's Analysis and Statistics group, and is the source of data on previous sanctions and proven reoffending.

The OASys research database.

Completed assessments are copied to the OASys Data Evaluation and Analysis Team (O-DEAT), a research and statistics office within NOMS headquarters. Data completeness and integrity checks are undertaken before producing subsets for analysis.

Participants

Offenders.

Two sets of OASys assessments were used. Assessments completed between January 2002 and September 2004 were used to construct the logistic regression model and develop OVP's scoring system. Assessments completed between October 2004 and September 2005 were accessed after OVP's content was finalised, and used as a formal validation sample.

All OASys assessments of offenders subject to Pre Sentence Reports, commencing community sentences or supervision upon release from custody were obtained from the O-DEAT database. At this point, the 2002-04 dataset contained 198,103 assessments, and the 2004-05 dataset contained 172,146 assessments. The assessments were systematically filtered to ensure complete data on dynamic risk factors and key matching variables, and to de-duplicate assessments relating to the same offender and sentence. Missing sentence details caused heavy attrition, especially in the 2002-04 period prior to improvements in data linkage with case management systems. Selection bias related to geographical area rather than offender characteristics such as age, gender, criminal history or dynamic risk. The remaining assessments were matched with the MoJPNC database on age, sex and index offence conviction date. Those successfully matched on age, sex and index offence conviction date were traced back from the conviction date to ascertain criminal history and traced forward from the sentence/release date to ascertain proven reoffending rates. The matched 2002-04 dataset comprised 15,918 assessments, and the matched 2004-05 dataset comprised 49,346 assessments. (A further 10,701 assessments from the 2002-04 dataset were initially used as a validation sample to meet NOMS's need for timely provisional results. These results are reported in Howard (2009), and are not reported here.) The 2002-04 dataset had a mean age of 29.7 years (SD 9.9 years), and included 14% female offenders, 7% of nonwhite ethnicity and a further 7% of unknown ethnicity. Thirty four percent had an index offence included in OVP's classification of violent offences (Howard & Dixon, 2011). The 2004-05 dataset had a mean age of 30.2 years (SD 10.1 years), and included 13% female offenders, 8% of nonwhite ethnicity and a further 9% of unknown ethnicity, and 35% with an OVP-class index offence.

OASys users.

Senior and main grade probation officers in four of NOMS's 42 probation areas participated in two pilot exercises during the development of OVP. Each officer scored OVP on several OASys-assessed offenders under their supervision. They later completed a questionnaire and/or participated in a focus group, seeking their views on OVP's scoring system and its utility in their assessment and case management practice. A central steering group met before and after each pilot and at key points in the subsequent implementation process. Its membership included managers from the pilot areas, policymakers, IT managers and researchers from NOMS HQ, and representatives of associated government agencies (e.g., the Parole Board).

Analysis

Overview.

An ordinal logistic regression model was fitted to predict proven violent reoffending in the 2002-04 construction dataset. The results of this model were manipulated so that each significant risk factor became a component of a weighted 100-point scale; a further 100-point scale with rounded factor weights and a unit-weighted scale were also created. The predictive validity of the original model and these three new scales were calculated for the 2004-05 validation dataset, enabling a decision on which of the four scoring options achieves the best balance of predictive validity and user friendliness. The process of generating predicted reoffending probabilities for fixed follow up durations from the selected rounded 100-point scale (the OVP score) is described. The predictive validity of the OVP score was compared with that of its dynamic and static subscales and other risk assessment tools, and checked for male/female and previously violent/nonviolent subgroups, again using the 2004-05 dataset.

Ordinal logistic regression modelling.

An ordinal logistic regression model was fitted on the 2002-04 construction sample (N=15,918). Ordinal logistic regression models produce predicted probabilities in a similar way to standard (binary) logistic regression models, but allow multiple outcomes which can be meaningfully ordered: in this instance, ranging from very rapid recidivism to less rapid recidivism to no recidivism. The dependent variable had 25 outcome categories: proven violent reoffending within 1, 2,.., 24 months, and no proven violent reoffending within a 24 month followup period. The model was fitted by a forward stepwise procedure with p= .05.

Four static risk factors were initially included: the number of previous sanctions for violent and nonviolent offences (including the index offence), age and gender. Age and previous sanction groups were created by successively dividing the distributions into increasing numbers of groups until further divisions failed to improve model fit or irregular patterns suggested a danger of overfitting the model. A fifth, binary, static variable for 'any previous (known) criminal history' status was created to reflect the substantially lower odds of proven reoffending of those with no known criminal history before the index offence.

The 32 dynamic factors were included: the ten OASys section scores, Moore's (2009) within-section and 'violence' factors, and other OASys questions which were not previously scored, including binary indicators of characteristics of the current offence, domestic violence perpetration and mental health problems. Some scales were disaggregated, following the observation by Mills, Kroner and Hemmati (2007) that the inclusion of individual items with limited associations with reoffending can damage model performance. To avoid overfitting, interactions were only modelled where a strong theoretical case for their existence was presented; none of these interactions proved significant, and they are not discussed further. Correlations between risk factors in the selected model confirmed that multicollinearity was absent. The ratio of events (i.e., reoffenders) to risk factors considered for model inclusion exceeded 40, whereas Harrell, Lee and Mark (1996) recommend a ratio of at least 10 and preferably over 20 to achieve adequate statistical power.

Simplification of the model to a 100-point scale.

To transform the logistic regression parameters into scores out of 100, the minimum and maximum possible scores based on the logistic regression parameters results were calculated. The range on each risk factor was expressed as hundredths of the overall range between the minimum and maximum. For example, the overall range of logistic regression parameters was 8.33 (from -3.14 to 5.19) and the range for accommodation was 0.17. Therefore, 2 points (100*0.17/8.33) were available for accommodation. Some small changes to these scores were necessary to overcome rounding effects and obtain a total of 100. Constant terms for a range of possible follow up periods were then estimated by re-entering the score into a further ordinal logistic regression model with 25 outcome categories, as in the previous step.

This scale was tested by OASys users in the first pilot. These staff criticised the resulting uneven distribution of item weights as unwieldy. Also, the low total weight of dynamic factors (26 of the 100 points) undermined the user requirement that the score should reflect an individual's progress over the course of their sentence. To deal with these two problems, a revision simplified the weighting system. Static factors were weighted at multiples of 5 and dynamic factors at multiples of 2, which rounded the weights considerably without causing too much disruption to the original model. The total weight for dynamic factors was raised from 26 to 40. (Simplified algorithms can replace regression results when combining scales into easily usable overall scores: Steyerberg, 2009.) OASys users who trialled this version of the scoring system in the second pilot found it much easier to comprehend. A still simpler unit weighted version was also created, reducing all risk factor weights to 0/1/2 or 0/2. A robust nonparametric test for correlated measures (DeLong, DeLong & Clarke-Pearson, 1988) was used to compare AUC predictive validity statistics for the original logistic regression model, original and revised 100-point score, unit weighted score, and the 60-point static and 40-point dynamic subscales of the revised score, for the validation sample (N=49,346) in SAS software version 9.2. Associations between the four scoring algorithms were measured with Spearman's rank correlation, in order to understand any differences in (rank-based) AUC statistics.

The revised 100-point scale was selected as the OVP score (see Results). Feedback from both pilots confirmed that users understood the distinction between the 100-point score and the 12- and 24-month probabilities, which are now routinely and automatically calculated on OASys summary sheets. (Predictions for the other 22 followup periods may be incorporated into parole procedures.) Examples are presented of predicted 24-month rates of violent reoffending and also homicide/ assault and homicide/ wounding reoffending, produced using a similar model, across a range of OVP scores. The OVP scores of violent reoffenders and nonreoffenders are illustrated.

Comparisons with other general and violence risk assessment tools.

The OGRS3 and RM2000/V scores were calculated from PNC data. The total OASys score was already included in the OASys assessments. RM2000/V is presented both in its usual categorised form, which takes four values between Low and Very High, and also the 9-point uncategorised score, to check whether any differences in predictive validity between OVP and RM2000/V are primarily due to the restricted range of the categorised version of RM2000/V. The static and dynamic subscales of OVP are also presented.

Utilising the validation sample, Area Under Curve scores were produced for three 24-month proven reoffending outcomes: any violent offence (i.e., any offence included in the OVP classification); homicide and assault offences (a consensus group, classified as violent by all existing violence risk assessment instruments (Howard and Dixon, 2011)), and homicide and wounding (the most serious offences within the consensus group). The DeLong et al (1988) test for significance of differences between AUCs determined whether the OVP score improved upon the predictive validity of the other tools for each of the three outcomes.

To illustrate the practical effects of such improvements in predictive validity, the distribution of OVP was equalised onto that of RM2000/V - the most predictive of the other tools – to eliminate OVP's longer range and thus permit fair comparison. As well as AUCs, sensitivity and specificity statistics were calculated to contrast actual and predicted reoffending when RM2000/V categories were used as the basis of three possible decision thresholds. (Similarly, OVP scores are reported as Low, Medium, High or Very High within the OASys summary sheet and RoSH.) When presenting these results, the size of the validation dataset was standardised to 10,000, for ease of comparison and generalization.

Comparisons of offender subgroups.

AUCs were calculated in order to check the predictive validity of the OVP score for four subgroups of offenders: females, males, those with no known history of violent offending and those with known history of such offending.

Results

Ordinal Logistic Regression Modelling

Table 1 shows the initial logistic regression model of proven violent reoffending. It selects five static risk items and seven dynamic risk items. The static risk factors selected include gender, a ten-point categorisation of age, a nine-point categorisation of violent criminal history and a four-point categorisation of nonviolent criminal history. Further discrimination between offenders with limited criminal careers is achieved through the 'any previous criminal history' item. The dynamic risk factors selected include two socioeconomic scales, alcohol misuse, two single items related to mental health and cognition, one item on attitudes towards the current offence, and one scale relating to attitudes towards crime, society and reoffending. OASys dynamic risk sections which did not contribute to the selected model were financial management and income, relationships, lifestyle and associates, and drug misuse.

TABLE 1 ABOUT HERE

Simplification of the Model to a 100-point Scale

Table 2 displays the minima, maxima and ranges of each coefficient of the logistic regression model, and the transformation of these ranges into scores on the original 100-point scale, and the revised 100-point and unit weighted scales. Table 3 gives further details of the scoring of the revised 100-point scale. In both 100-point scales, previous violent sanctions and age carried the highest weights, while alcohol misuse was the highest weighted dynamic factor. Some disproportionate changes occurred due to the simplification of weights and the transference of 14 points from static to dynamic risk factors for the revised scale. The weights for nonviolent sanctions and the 'any previous criminal history' item were reduced sharply, while the weights for recognising the impact of offending and accommodation were doubled.

TABLES 2 AND 3 ABOUT HERE

Table 4 shows that validation sample AUCs for the logistic regression model and the two 100-point scales differed with neither clinical nor statistical significance for homicide/ wounding reoffending. While comparisons of the logistic regression model with the revised score were statistically significant for both homicide/assault ($\chi^2 = 6.14$, $p = .013$) and all violent reoffending ($\chi^2 = 37.80$, $p < .001$), but the AUC differences remained no greater than .003 and thus had little clinical significance. The unit weighting produced a considerably lower AUC for all three outcome measures. Spearman's correlations between the logistic regression scores and other scores were .99 (original 100-point), .98 (revised 100-point) and .88 (unit weighted). This shows that the original and revised 100-point scales had minimal impacts on the rank orders of offenders' scores and thus on AUCs, whereas applying unit weighting had an appreciable impact.

On the basis of these results and the views of OASys users, the revised 100-point scale was selected, and named the OASys Violence Predictor (OVP).

TABLE 4 ABOUT HERE

Table 5 presents predicted 24-month proven reoffending probabilities for a range of OVP scores, covering homicide/wounding, homicide/assault and all violent reoffences. The fitted probabilities for homicide/wounding indicate the concentration of this outcome among those with the highest OVP scores. Figure 1 shows that 100-point scores among 24-month reoffenders and nonreoffenders for all violent offences were both approximately normally distributed, with mean 47.7 and standard deviation 12.0 among reoffenders and mean 36.0 and standard deviation 12.9 among nonreoffenders.

TABLE 5 ABOUT HERE

FIGURE 1 ABOUT HERE

Table 6 compares the predictive validity of OVP and the other risk assessment tools. Rice and Harris (2005) suggest that AUCs of .71 or higher are good, while AUCs of .64 or higher are moderate. OVP achieved AUCs of .74 for all violent offences, compared with .66-.70 for the four other tools. The intermediate outcome of homicide/ assault was most difficult to predict, with AUCs of .71 for OVP and .63-.68 for other tools. The most serious outcome, homicide/wounding, was predicted with AUCs of .72 for OVP and .65-.68 for other tools. All comparisons between OVP and other tools were significant at $p < .001$. RM2000/V performed better as a 9-point raw score (0-8) than in four categories. When OVP was synchronised with the four RM2000/V categories, thus eliminating its range advantage, this four-category OVP had AUCs of .699 (.677, .721), .690 (.684, .695) and .719 (.714, .723) for homicide/wounding, homicide/assault and all violent reoffending respectively. The paired comparison test with the four-category RM2000/V returned *p* values of .0109, <.001 and <.001 respectively. Taken together, these results show that OVP's superior predictive validity was partially due to distributional effects (i.e., its longer range), but partially – and still statistically and clinically significantly – due to better discrimination between reoffenders and nonreoffenders after controlling for such effects.

The static subscale of OVP was a better predictor than RM2000/V, with AUCs of .70-.73. It was a significantly weaker predictor than the total OVP score for homicide/assault and all OVP offences ($p < .001$), but not for homicide/wounding ($p = .10$). The dynamic subscale was moderately predictive, with AUCs of .65-.67 for the three outcomes.

TABLE 6 ABOUT HERE

Table 7 compares the sensitivity and specificity of OVP and RM2000/V, standardising OVP's distribution onto the RM2000/V categorisation as described above. In a caseload of 10,000, using OVP rather than RM2000/V results in an identically-sized Very High risk category being used for an additional 9 of the 90 homicide/wounding reoffenders, 38 of the remaining 1,502 less serious assault reoffenders, and 79 of the remaining 1,158 reoffenders convicted of other OVP-class violent (i.e., robbery, aggravated burglary, threats/harassment, weapon possession, public order and/or criminal damage) reoffences. OVP's Low risk category would include 28 less serious serious assault reoffenders compared with RM2000/V's 43, and 16 rather than 38 other violent reoffenders. (Both predictors' Low risk categories included one homicide/wounding reoffender.) As the Discussion below details, this resource-neutral categorisation shift evidently gives corrections staff many additional opportunities to intervene to prevent harmful reoffending.

TABLE 7 ABOUT HERE

Comparisons of Offender Subgroups

Table 8 reports the AUCs and confidence intervals of OVP as a predictor of the three outcomes among the four offender subgroups, as well as the rates of these outcomes. Most of these AUCs are somewhat lower than those in Table 4. This was expected, as the subgroups were homogeneous on particular risk factors for violent reoffending (i.e., gender or violence history) and therefore spanned narrower ranges of relative risk (Hanson, 2008). All three outcomes had higher AUCs among female than male offenders; paired comparison tests returned *p* values of <.001 for the two more frequent outcomes, but were not significant for homicide/wounding. The difference between female and male rates was far greater for homicide/wounding (OR 0.32) than homicide and assault or all violent offending (ORs 0.68 and 0.56 respectively). AUCs were higher among those with no known history of violence than those with known history, with *p* values of <.001 for the two more frequent outcomes and no significant difference for homicide/wounding. Reoffending rates were far lower among those with no history of violence, but were not negligible; these offenders comprised 26% of the sample, 11% of violent and homicide/assault reoffenders and 8% of homicide/wounding reoffenders.

TABLE 8 ABOUT HERE

Discussion

Prediction of violent recidivism is a key activity for large correctional organisations such as NOMS. Information on likely future harmful offending is vital if scarce intervention and supervision resources are to be allocated efficiently. However, resource constraints prevent the routine use of risk assessment instruments that can only be administered by trained mental health professionals. This study developed a violence risk predictor which can be scored and used to identify higher-risk offenders within NOMS's principal risk assessment and management system, OASys. Resource allocation is strongly influenced by OASys risk of serious harm ratings, of which OVP scores are a major determinant. Offenders identified as higher-risk receive more intensive supervision, supervision by senior staff members, and prioritisation for treatment on accredited groupwork programmes (National Offender Management Service, 2010). They may also be subject to enhanced management under Multi Agency Public Protection Arrangements (Ministry of Justice, 2010b).

Improving Predictive Validity

An ordinal logistic regression model was fitted, utilising a very large sample of OASys assessments matched with official offending data. It selected static risk factors encompassing age, gender, and general and violent criminality domains, and dynamic risk factors encompassing socioeconomic, substance misuse, mental health, cognitive and attitudinal domains. Considerable simplification and rounding of the model coefficients created a more user-friendly scoring system while sacrificing very little predictive validity. A further simplification to unit weighting would have incurred a far greater loss of validity and was rejected. The resulting OASys Violence Predictor substantially improved predictive validity for nonsexual violent recidivism, whether defined narrowly or broadly, compared with predictors of general and nonsexual violent recidivism routinely available in NOMS. It also demonstrated moderate to high levels of predictive validity within offender subgroups. OVP has now been implemented in an upgrade of OASys, which revised its criminal history section and summary sheet. The summary sheet's IT functionality improves OVP's field validity by automatically calculating predictor scores. A separate predictor of nonviolent reoffending, the OASys General reoffending Predictor (OGP), was introduced simultaneously, and the existing "OASys score" withdrawn (Howard, 2009). OGP is structured in the same way as OVP, but with differences in composition including the use of OGRS3 as the static risk component and the presence of drug misuse and absence of alcohol misuse among the dynamic risk factors.

The difference in predictive validity between OVP and the other tools runs contrary to some recent conclusions. While OVP is the only viable static/dynamic predictor for widespread use in NOMS, it is feasible for other predictors to be used for offenders serving long or indeterminate prison sentences. For these offenders, further studies should check whether OVP improves upon the validity of existing violence risk predictors and/or can be beneficially used in combination with them i.e., achieves incremental validity. Such studies would ideally directly capture complete scores or SPJ risk judgments on each major instrument, and be sufficiently large to detect moderate differences (e.g., two to four points of AUC) in effect sizes. Conversely, OASys is not routinely used in forensic mental health settings or outside England and Wales. It is not certain that OVP's advantage would be maintained in these other settings. Patterns of offending behavior may vary, or criminal justice systems could operate differently (e.g., affecting the size of the gap between true and officially recorded offending), affecting the validity of dynamic and static risk factors respectively. It may be feasible for other European jurisdictions using risk assessment systems based on OASys (van Durnescu & Kalmthout, 2008) to calculate OVP scores from existing research datasets and thereby compare it with any other predictors in routine use.

The prospects for OVP to demonstrate strong predictive validity in further studies seem reasonable. It combines a strong focus on relevant static factors – principally age and previous violence - with a range of clinically relevant dynamic risk factors in a novel manner: “Although actuarial guides could, in theory, include causal dynamic risk factors, extant measures heavily weight static variables, nearly to the exclusion of dynamic ones” (Douglas & Skeem, 2005, p. 352). Of the other tools in the current study, OGRS3 and RM2000/V fail to cover all relevant static factors, omitting previous violence and gender respectively. Further validation studies would be required to allow comparison with other instruments. Of the existing tools, LSI-R and LS/CMI are most similar to OASys and have performed well in recent meta-analyses. We suggest that their prediction of violent recidivism could improve somewhat if selected items were used to create a violence-specific scale, in the same way as OVP selects specific OASys items which are relevant to violence risk. As a more general point, future research is unlikely to identify radical changes in prediction methodology which can produce and sustain massive improvements in predictive validity. Rather, making technical efficiencies (such as OVP’s empirically-based violence classification) and combining the best features of existing approaches can lead to incremental improvements which appear modest on a case-by-case basis but provide substantial real-world benefit when applied to large correctional caseloads. It may be more difficult to prove the value of such approaches in organisations with smaller caseloads. .

Assessment of Dynamic Risk Factors

The full static/dynamic OVP score was only slightly more predictive than the static OVP score, despite the moderate association between the dynamic OVP score and recidivism. The limited improvement in predictive validity attributable to dynamic risk factors is consistent with recent research on NOMS offenders (Yang, Liu & Coid, 2010). In OASys, dynamic risk assessment is not only useful as a source of predictive validity, with the benefits in case management and resource allocation outlined above. Risk factors identified in OASys are also key indicators of need and responsivity when setting sentence plans and allocating treatment (National Offender Management Service, 2010). In settings where assessments are used for risk prediction alone, the predictive benefits of dynamic factors would have to exceed the costs of assessing them.

While our results might imply pessimism about the predictive value of dynamic risk factors, our method follows the common research assumption that risk is assessed only once, at the start of community supervision. This creates an inbuilt bias against genuinely dynamic risk factors, which must be reassessed periodically to remain relevant. Recent research (Brown et al, 2009; Jones, Brown & Zamble, 2010; Olver, Wong, Nicolaichuk & Gordon, 2007) has favourably examined the predictive utility of dynamic factors in the more realistic context of repeated assessment. Further research (Howard and Dixon, 2011) has examined the dynamic elements of OVP in the context of NOMS practice, where OASys assessments are reviewed every few months. This has determined that most of the risk factors in OVP's dynamic element do have the properties of causal dynamic risk factors, with changes in the probability of violent recidivism occurring after changes in risk factor scores (Kraemer et al, 1997). Rigorously testing these qualities in OVP's risk factors addresses a topic of considerable research interest. In NOMS practice, it provides evidence that OASys identifies valid treatment targets, and that reassessment during ongoing correctional supervision is a worthwhile activity. Changes in causal dynamic risk factors could potentially be measured in treatment evaluations, both as interim outcomes before reoffending data become available and later to help explain the intervention's success or failure in reducing reoffending.

Methodological Considerations

This study benefitted from an extremely large sample size, which provided narrow confidence intervals for parameter estimates. It was therefore possible to apply nonunit weights while avoiding shrinkage (loss of model fit), contrary to the unit weighting recommendations of earlier, smaller, studies (e.g., Grann & Langstrom, 2007, where *N*=404). Item weights were however smoothed, creating a practitioner-friendly and fairly dynamic scoring system. This incurred only minimal loss of predictive validity.

OVP's predictive validity might be improved further if OASys included a broader range of (putatively) dynamic risk factors. The present psychiatric treatment item seems crude but reliable: unlike other items in section 10 (Emotional Well-being), only basic information and training are required to score it. Most OASys assessments record little or no direct information on personality disorder, psychopathic personality features or active psychotic symptoms. Likewise, the predictor's blanket rejection of drug misuse ignores the probable link between use of illegal stimulants (e.g., amphetamines, cocaine) and violence (Boles & Miotto, 2003). (OASys does record use of a range of different drugs, but data completeness was poor during the period in which our construction assessments were completed.)

The use of actuarially scored tools such as OVP is founded upon the principle that they produce meaningful predictions for individual subjects. However, two recent papers argue that the 'precision intervals' of these predictions are so wide as to be essentially meaningless (Hart, Cooke & Michie, 2007; Cooke & Michie, 2010). Several authors, most recently Hanson and Howard (2010), have criticised the statistical reasoning and conclusions of these papers. NOMS research managers concur with these criticisms, and are therefore content to use OVP and other actuarial tools. OVP and, for known sexual offenders, Risk Matrix 2000/S scores provide a firm objective basis for risk of serious harm ratings and therefore treatment and supervision intensity. Nevertheless, allowance is made for cautious clinical modifications on the basis of "human judgement and experience" (Gottfredson & Moriarty, 2006, p. 17).

Conclusion

In conclusion, the introduction of OVP represents a considerable improvement on the predictors routinely available in NOMS practice. The improvements in predictive validity detected and cross-validated here are modest on a per-case basis but create considerable public benefit when applied on the scale of a large correctional system. OVP provides valid information on violence risk for a very large group of offenders responsible for a substantial number of harmful reoffences. Its potential for application in forensic mental health settings, where risk assessment tools such as HCR-20, VRAG and PCL-R are available, should be considered carefully, and the approach taken could be replicated in other jurisdictions using comprehensive risk/need assessment systems. The extent to which changes in OVP's dynamic risk scores over time aid prediction of reoffending will be examined in a forthcoming study.

References

Andrews, D.A., & Bonta, J. (1995). *Level of Service Inventory – Revised*. Toronto: Multi-Health Systems.

Andrews, D.A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.) Cincinnati: Anderson Publishing.

Andrews, D.A., Bonta, J., & Wormith, J.S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto: Multi-Health Systems.

Boles, S.M., & Miotto, K. (2003). Substance abuse and violence: A review of the literature. *Aggression and Violent Behavior, 8*, 155-174. doi 10.1016/S1359-1789(01)00057-X

Brown, S.L., St. Amand, M.D., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law & Human Behavior, 33*, 25-45. doi 10.1007/s10979-008-9139-7

Campbell, M.A., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism among adult offenders*. (Public Safety Canada Research Report, 2007-04.) Retrieved from Public Safety Canada website: http://www.ps-sp.gc.ca/res/cor/rep/risk-tools-200704-eng.aspx

Cooke, D.J., & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior, 34,* 259-274. doi: 10.1007/s10979-009-9176-x

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated ROC curves: A nonparametric approach. *Biometrics, 44*, 837-845.

Douglas, K.S., & Skeem, J.L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383. doi: 10.1037/1076-8971.11.3.347

Douglas, K.S., Yeomans, M., & Boer, D.P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*, 479-510. doi 10.1177/0093854805278411

Edens, J.F., & Douglas, K.S. (2006). Assessment of interpersonal aggression and violence: Introduction to the special issue. *Assessment, 13*, 221-226. doi:0.1177/1073191106291443

Gottfredson, S.D., & Moriarty, L. J. (2006). Clinical versus actuarial judgments in criminal justice decisions: should one replace the other? *Federal Probation, 70 (2)*, 15-18.

Grann, M., & Langstrom, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior, 34*, 22-36. doi: 10.1177/0093854806290250

Hanson, R.K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence, 20*, 212-217. doi: 10.1177/0886260504267740

Hanson, R.K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene, 15 (1)*, 15-17. Retrieved from Canadian Psychological Association website: http://www.cpa.ca/cpasite/userfiles/Documents/Criminal%20Justice/Crime%20Scene%202008-04.pdf

Hanson, R.K., & Howard, P.D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34,* 275-281. doi: 10.1007/s10979-010-9227-3

Hare, R.D. (2003). *The Hare Psychopathy Checklist – Revised* (2nd ed.). Toronto: Multi-Health Systems.

Harrell, F.E., Lee, K.L., & Mark, D.B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine, 15*, 361-387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Hart, S.D., Cooke, D.J., & Michie, C. (2007). Precision of actuarial risk assessment instruments: Evaluating the "margins of error" of group v. individual predictions of violence. *British Journal of Psychiatry, 190*, 60-65. doi:10.1192/bjp.190.5.s60

Heilbrun, K., Yasuhara, K., & Shah, S. (2009). Violence risk assessment tools: Overview and clinical analysis. In Otto, R.K., & Douglas, K.S. (Eds.), *Handbook of violence risk assessment* (pp. 1-18). New York: Routledge.

HM Prison Service (2009). Generic parole process. (Prison Service Order 6010). Retrieved from http://www.pso.hmprisonservice.gov.uk/pso_6010_generic_parole_process.doc

HM Treasury (2010). *Spending Review 2010*. London: HM Treasury.

Home Office (2006). *Offender Assessment System OASys User Manual. (Revised July 2006.)* London: Home Office.

Howard, P. (2006). *The Offender Assessment System: An evaluation of the second pilot* (Home Office Research Findings 278). Retrieved from Home Office website: http://www.homeoffice.gov.uk/rds/pdfs06/r278.pdf

Howard, P. (2009). *Improving the prediction of reoffending using the Offender Assessment System* (Ministry of Justice Research Summary 2/09). Retrieved from Ministry of Justice website: http://www.justice.gov.uk/publications/docs/oasys-research-summary-02-09.pdf

Howard, P.D., & Dixon, L. (2011). Developing an empirical classification of violent offenses for use in the prediction of recidivism in England and Wales. *Journal of Aggression, Conflict and Peace Research, 3*, 141-154. doi: 10.1108/17596591111154176

Howard, P.D., & Dixon, L. (2011). *Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys Violence Predictor*. Manuscript submitted for publication.

Howard, P., Francis, B., Soothill, K., & Humphreys L. (2009). *OGRS 3: The revised Offender Group Reconviction Scale* (Ministry of Justice Research Summary, 7/09). Retrieved from Ministry of Justice website: http://www.justice.gov.uk/publications/docs/oasys-research-summary-07-09-ii.pdf

Howard, P., & Moore, R. (2009). *Measuring changes in risk and need over time using OASys* (Ministry of Justice Research Summary, 10/09). Retrieved from Ministry of Justice website: http://www.justice.gov.uk/publications/docs/research-summary-oasys-10-09.pdf

Howells, K. (2009). Recent development and future directions in understanding cognition and emotion in offenders: a commentary. *Psychology, Crime & Law, 15*, 285-292. doi: 10.1080/10683160802191036

Jones, N.J., Brown, S.L, & Zamble, E. (2010). Predicting criminal recidivism in adult male offenders: Researcher versus parole officer assessment of dynamic risk. *Criminal Justice and Behavior, 37*, 860-882. doi: 10.1177/0093854810368924

Kraemer, H.C., Kazdin, A.E., Offord, D.R., Kessler, R.C., Jensen, P.S., & Kupfer, D.J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337-343.

Kroner, D.G., Mills, J.F., & Reddon, J.R. (2005). A Coffee Can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry, 28*, 360-374. doi: 10.1016/j.ijlp.2004.01.011

Mair, G., Burke, L.,& Taylor, S. (2006). 'The worst tax form you've ever seen'? Probation officers' views about OASys. *Probation Journal, 53*, 7-23. doi: 10.117710264550506060861

Mills, J.F., Kroner, D.G., & Hemmati, T. (2007). The validity of violence risk assessments: An issue of individual item performance. *Psychological Services, 4*, 1-12. doi: 10.1037/1541-1559.4.1.1

Ministry of Justice (2006). *Implementation of the risk of harm guidance and training resource pack*. (Probation Circular 22/2006.) Retrieved from Home Office website: http://www.probation.homeoffice.gov.uk/files/pdf/PC22%202006.pdf

Ministry of Justice (2010a). *Offender caseload management statistics 2009*. Retrieved from http://www.justice.gov.uk/publications/prisonandprobation.htm

Ministry of Justice (2010b). *Multi-Agency Public Protection Arrangements Annual Report 2009/10.* Retrieved from http://www.justice.gov.uk/publications/docs/mappa-stats-release.pdf

Moore, R. (2009). *The internal reliability and construct validity of the Offender Assessment System (OASys)* (Ministry of Justice Research Summary 6/09). Retrieved from Ministry of Justice website: http://www.justice.gov.uk/publications/docs/oasys-research-summary-06-09.pdf

Morton, S. (2009). *Can OASys deliver consistent assessments of offenders? Results from the inter-rater reliability study* (Ministry of Justice Research Summary 1/09). Retrieved from Ministry of Justice website: http://www.justice.gov.uk/docs/oasys-research-summary-01-09.pdf

National Offender Management Service (2010). *Suitability for accredited interventions: Risk, need and responsivity targeting criteria for offender managers and treatment managers and HMPS therapy managers*. Unpublished internal government document.

National Probation Service (2008). *National rules for tiering cases and associated guidance* (Probation Circular 8/08). Retrieved from http://www.probation.homeoffice.gov.uk/files/pdf/PC08%202008.pdf

National Probation Service (2009). *Determining pre-sentence report type*. (Probation Circular 6/09.) Retrieved from http://www.probation.homeoffice.gov.uk/files/pdf/PC06%202009.pdf

Olver, M.E., Wong, S.C.P., Nicolaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale – Sexual Offender version. Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318-329. doi: 10.1037/1040-

3590.19.3.318

Palmer, E.J, McGuire, J., Hatcher, R.M., Hounsome, J.C., Bilby, C.A.L., & Hollin, C.R. (2009). Allocation to offending behavior programs in the English and Welsh probation service. *Criminal Justice and Behavior, 36*, 909-922. doi: 10.1177/0093854809338889

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior, 29*, 615-620. doi: 10.1007/s10979-005-6832-7

Seto, M.C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment, 17*, 156-167. doi: 10.1037/1040-3590.17.2.156

Steyerberg, E.W., (2009). *Clinical prediction models: A practical approach to development, validation and updating*. New York: Springer.

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: a Journal of Research and Treatment, 14*, 139-154. doi: 10.1177/107906320201400205

Thornton, D. (2007). *Scoring Guide for Risk Matrix 2000.9/SVC.* Retrieved from http://www.cfcp.bham.ac.uk/Extras/SCORING%20GUIDE%20FOR%20RISK%20MATRIX%202000.9-%20SVC%20-%20(ver.%20Feb%202007).pdf

van Kalmthout, A.M, & Durnescu, I. (Eds.) (2008). *Probation in Europe*. Oisterwijk, The Netherlands: Wolf Legal Publishers.

Walters, G.D. (2007). Methodological and statistical advances in correctional and forensic psychology. *Criminal Justice and Behavior, 34*, 1533-1535. doi: 10.1177/0093854807307024

Webster, C.D., Douglas, K.S., Eaves, D. & Hart, S.D. (1997). *HCR-20. Assessing Risk for Violence. Version 2.* Vancouver, Canada: Simon Fraser University.

Yang, M., Liu, Y., & Coid, J. (2010). *Applying neural networks and other statistical models to the classification of serious offenders and the prediction of recidivism* (Ministry of Justice Research Series 6/10). Retrieved from Ministry of Justice website: http://www.justice.gov.uk/publications/docs/neural-networks-research.pdf

Yang, M., Wong, S.C.P., & Coid, J. (2010). The efficacy of violence prediction. A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767. doi: 10.1037/a0020473

Table 1

Logistic regression model of proven violent reoffending within 12 and 24 months (N=15,918)

Variable	*B*	SE *B*	Odds ratio
Constant (12 months)	-2.00	0.096	--
Constant (24 months)	-1.35	0.096	--
Number of sanctioning occasions for violent-type offences			
None	-1.09	0.08	0.09
1	-0.79	0.07	0.12
2	-0.53	0.07	0.15
3	-0.39	0.08	0.17
4	-0.25	0.08	0.20
5	-0.22	0.10	0.21
6	-0.07	0.11	0.24
7 or 8	0.12	0.11	0.29
9 or 10	0.40	0.15	0.38
11 or 12	0.68	0.22	0.51
13 to 17	0.77	0.26	0.55
18 or more (reference category)	1.37	--	1
Number of sanctioning occasions for other offences			
None, 1 or 2	-0.37	0.04	0.49
3 or 4	-0.14	0.04	0.62
5 to 10	-0.05	0.04	0.68
11 to 20	0.20	0.04	0.87
21 or more (reference category)	0.36	--	1
Does the offender have any previous sanctions?			
Yes	0.32	0.04	1.88
No	-0.32	--	--
Age at date of assessment, grouped			
18-19	0.97	0.05	6.91
20-21	0.68	0.05	5.13
22-23	0.45	0.06	4.08
24-25	0.40	0.06	3.89
26-30	0.19	0.05	3.16
31-35	-0.13	0.05	2.29
36-40	-0.37	0.06	1.80
41-45	-0.58	0.08	1.47
46-50	-0.65	0.12	1.37
51+ (reference category)	-0.96	--	1
Sex			
Female	-0.22	0.04	0.64
Male (reference category)	0.22	--	1

2.6: Recognizes impact of current offending on victim, community and wider society?			
Yes	-0.08	0.03	1.18
No (reference category)	0.08	--	1
3.3 to 3.6: Score on four accommodation questions (range 0-8)	0.02	0.01	1.02 / 1.18
4.2 to 4.5: Score on four employment questions (range 0-8)	0.04	0.01	1.05 / 1.42
9.1, 9.2: Current [chronic] alcohol misuse and binge drinking (range 0-4)	0.15	0.01	1.16 / 1.80
10.7: Current psychiatric treatment, or treatment pending			
Yes	0.11	0.04	1.25
No (reference category)	-0.11	--	1
11.4: Temper control (range 0-2)	0.19	0.03	1.21 / 1.46
12.1, 12.3 to 12.8: Score on six attitudes questions (range 0-12)	0.02	0.01	1.02 / 1.30

Note. Number(s) preceding variable descriptions denote OASys question number(s). Where two odds ratios are given, the first compares a score of 1 with a reference score of 0, and the second compares the maximum score on this variable's range with a reference score of 0. Question wording and scoring options (0/1/2 unless stated): 3.3, Currently of no fixed abode or in transient accommodation (0/2; if 2, score 3.4, 3.5 and 3.6 as 2 also); 3.4, Suitability of accommodation; 3.5, Permanence of accommodation; 3.6, Suitability of location of accommodation; 4.2, Unemployed, or will be unemployed on release (0/2); 4.3, Employment history; 4.4, Work-related skills; 4.5, Attitude to employment; 9.1, Is current alcohol use a problem; 9.2, Binge drinking or excessive use of alcohol in last 6 months; 12.1, Pro-criminal attitudes; 12.3, Attitude towards staff; 12.4, Attitude towards supervision/licence; 12.5, Attitude towards community/society; 12.6 Does the offender understand their motivation for offending; 12.8 Is the offender motivated to address offending.

Table 2
Scaling and adjustment of logistic regression results to produce three simplified scores

Risk factor	Raw regression parameters			Risk factor weights in simplified models		
	Minimum	Maximum	Range	Original	Revised	Unit weights
Sanctions for violent offences	-1.09	1.36	2.45	29	25	2
Sanctions for other offences	-0.36	0.36	0.72	9	5	2
Any previous sanctions	-0.32	0.32	0.63	8	5	2
Age	-0.96	0.96	1.92	23	20	2
Sex	-0.22	0.22	0.44	5	5	2
All static factors	*-2.95*	*3.22*	*6.17*	*74*	*60*	*10*
2.6: Recognizes impact of offending	-0.08	0.08	0.16	2	4	2
3.3-3.6: Accommodation	0	0.17	0.17	2	4	2
4.2-4.5: Employability	0	0.36	0.36	4	6	2
9.1-9.2: Alcohol misuse	0	0.60	0.60	7	10	2
10.7: Psychiatric treatment	-0.11	0.11	0.22	3	4	2
11.4: Temper control	0	0.40	0.40	5	6	2
12.1, 12.3-12.8: Attitudes	0	0.25	0.25	3	4	2
All dynamic factors	*-0.19*	*1.97*	*2.16*	*26*	*40*	*14*
Total	*-3.14*	*5.19*	*8.33*	*100*	*100*	*24*

Note. The original model weights equal the range of the raw regression parameters multiplied by 100/8.33 and rounded. Scores on the original and unit weighted models were calculated as the rounded value of weight*(coefficient-minimum)/range. Detailed scoring of the revised model is reported in Table 3.

Table 3
Scoring of OVP from static and dynamic risk factors

Static risk factor	Weight	Dynamic risk factor	Weight
Sanctions for violent offences		Recognizes impact of current offence	
None	0	No	4
1	4	Yes	0
2	7	Accommodation score	
3	9	0	0
4	11	1 or 2	1
5	12	3 or 4	2
6	13	5 or 6	3
7	14	8	4
8	15	Employability score	
9	16	0	0
10	17	1	1
11	18	2 or 3	2
12	19	4	3
13	20	5	4
14	21	6 or 7	5
15	22	8	6
16	23	Alcohol misuse score	
17	24	0	0
18 or more	25	1	3
Sanctions for other offences		2	5
None, 1 or 2	0	3	8
3 or 4	2	4	10
5 to 10	3	Psychiatric treatment	
11 to 20	4	Not current or pending	0
21 or more	5	Current or pending	4
Previous sanctions		Temper control score	
Any previous sanctions	5	0	0
No previous sanctions	0	1	3
Age		2	6
18 or 19	20	Attitudes score	
20 or 21	17	0	0
22 or 23	14	1 or 2	1
24 or 25	12	3 or 4	2
26 to 30	10	5 or 6	3
31 to 35	8	7 or 8	4
36 to 40	6	9 or 10	5
41 to 45	4	11 or 12	6
46 to 50	2		
51 or older	0		
Sex			
Female	0		
Male	5		

Note. An accommodation score of 7 is not possible, as question 3.3 can only be scored 0 or 2, and if it is scored 2 then questions 3.4, 3.5 and 3.6 are also scored 2.
Predicted probability = e(z) / 1 + e(z), where z = -4.522 + (0.0722 * score) [12 months] or z = -3.877 + (0.0722 * score) [24 months].

Table 4

Predictive validity of raw logistic regression parameters and three simplified scores (N=49,346)

Model	AUC and 95% CIs for proven reoffending involving this offence type within 24 months (% reoffending)		
	Homicide & wounding (0.9%)	Homicide & assault (15.9%)	All violent offences (27.5%)
Raw logistic regression parameters	.726 (.704, .747)	.715 (.709, .720)	.748 (.744, .753)
Original score/100	.723 (.701, .745)	.715 (.709, .721)	.748 (.743, .752)
Revised score/100 ["OVP score"]	.723 (.701, .745)	.713 (.707, .719)	.745 (.740, .749)
Unit weighted score	.699 (.676, .722)	.686 (.680, .692)	.715 (.710, .720)

Note. T-test comparisons between raw regression parameters and simplified scores were significant at: * p<.05, ** p<.01, *** p<.001.

Table 5

Predicted probability of homicide & wounding, homicide & assault and all violent offences, for a range of OVP scores

OVP score	% proven reoffending involving this offence type within 24 months		
	Homicide & wounding (0.9%)	Homicide & assault (15.9%)	All violent offences (27.5%)
0	0.1%	1.5%	2.0%
10	0.1%	2.7%	4.1%
20	0.2%	4.7%	8.1%
30	0.4%	8.2%	15.3%
40	0.7%	13.9%	27.1%
50	1.2%	22.5%	43.4%
60	2.2%	34.4%	61.2%
70	3.9%	48.6%	76.4%
80	6.8%	63.1%	87.0%
90	11.6%	75.5%	93.2%

Note: predicted probability of all violent offences as Table 3. Predicted probability of homicide and assault: $z = -4.185 + (0.0590 * \text{score})$. Predicted probability of homicide and wounding: $z = -7.337 + (0.0590 * \text{score})$.

Table 6
Comparisons of predictive validity between OVP total score and other predictors (N=49,346)

Risk assessment score	AUC and 95% CIs for proven reoffending involving this offence type within 24 months (% reoffending)		
	Homicide & wounding (0.7%)	Homicide & assault (13.7%)	All violent offences (24.9%)
OVP total score	.723 (.701, .745)	.713 (.707, .719)	.745 (.740, .749)
OVP static score	.712 (.696, .734)	.700 (.694, .707)	.733 (.728, .737)
OVP dynamic score	.659 (.634, .684)	.652 (.645, .658)	.672 (.667, .677)
OASys score	.654 (.632, .678)	.634 (.627, .640)	.659 (.654, .664)
OGRS3 2-year percentage	.658 (.636, .680)	.665 (.659, .671)	.697 (.692, .702)
Risk Matrix 2000/V category	.667 (.648, .690)	.664 (.658, .670)	.674 (.669, .679)
Risk Matrix 2000/V score	.683 (.661, .704)	.680 (.674, .686)	.690 (.685, .695)

Note. All comparisons of predictive validity between the OVP total score and other predictors were significant at $p<.001$, except the comparison of OVP total and static score for homicide & wounding reoffending was not significant ($p=.10$).

Table 7
Sensitivity and specificity of OVP and Risk Matrix 2000/V categories

Risk categories / score range below cutoff	Risk categories / score range above cutoff	% of all offenders above cutoff	Predictor	Outcomes per 10,000 cases				Sensitivity	Specificity
				TP	FP	FN	TN		
Homicide & wounding (0.7% proven reoffending)									
Low	Medium, High, Very High	88.4	RMV	89	8752	1	1158	98.6	11.7
			OVP	89	8752	1	1158	98.9	11.7
Low, Medium	High, Very High	52.4	RMV	72	5168	18	4742	80.4	47.8
			OVP	73	5167	17	4743	81.3	47.9
Low, Medium, High	Very High	15.0	RMV	27	1468	63	8442	30.2	85.2
			OVP	36	1459	54	8451	40.2	85.3
Homicide & assault (15.9% proven reoffending)									
Low	Medium, High, Very High	88.4	RMV	1548	7292	44	1116	97.2	13.3
			OVP	1563	7277	29	1131	98.2	13.4
Low, Medium	High, Very High	52.4	RMV	1186	4054	406	4354	74.5	51.8
			OVP	1235	4005	357	4403	77.6	52.4
Low, Medium, High	Very High	15.0	RMV	456	1039	1136	7369	28.7	87.6
			OVP	503	992	1089	7416	31.6	88.2
All violent offences (27.5% proven reoffending)									
Low	Medium, High, Very High	88.4	RMV	2668	6173	82	1078	97.0	14.9
			OVP	2705	6135	45	1115	98.4	15.4
Low, Medium	High, Very High	52.4	RMV	2003	3238	747	4013	72.8	55.3
			OVP	2132	3108	617	4142	77.5	57.1
Low, Medium, High	Very High	15.0	RMV	730	765	2020	6485	26.5	89.4
			OVP	856	639	1893	6612	31.1	91.2

Note. RMV = Risk Matrix 2000/V. For the purpose of comparison, OVP score ranges were selected to match the distribution of Risk Matrix 2000/V categories, with ties broken randomly (0-22 = Low; 22-38 = Medium; 38-54 = High; 54-100 = Very High). These are not the score ranges used in NOMS practice. “Per 10,000 cases” results represent a 10,000-strong caseload with identical reoffending rates and score distributions to the original 49,436 cases. Sensitivity and specificity are calculated from the original 49,436 cases, rather than the simulated 10,000 cases. TP = true positive; FN = false negative; FP = false positive; TN = true negative.

Table 8
Comparisons of the predictive validity of OVP between offender subgroups (N=49,346)

Offender subgroup *(n)*	AUC, 95% CIs and rates of proven reoffending involving this offence type within 24 months					
	Homicide & wounding (0.7%)		Homicide & assault (13.7%)		All violent offences (24.9%)	
	% reoffending	AUC (CI)	% reoffending	AUC (CI)	% reoffending	AUC (CI)
Female *(6,368)*	0.31%	.733 (.629, .836)	11.5%	.741 (.723, .759)	18.3%	.765 (.751, .780)
Male *(42,978)*	0.98%	.711 (.687, .734)	16.2%	.707 (.700, .713)	28.6%	.738 (.733, .742)
No sanctions for violent offences *(12,671)*	0.26%	.711 (.634, .789)	6.7%	.706 (.688, .723)	12.1%	.739 (.726, .752)
Sanctions for violent offences *(36,675)*	1.12%	.683 (.658, .708)	18.7%	.673 (.666, .680)	32.7%	.704 (.698, .710)